

Graphs, Harry Potter, Cytoscape, and

David K. Kloster
Cyberinfrastructure for Digital Humanities
Indiana University
October 20, 2017

Cyberinfrastructure for Digital Humanities

Indiana University

October 20, 2017



University Information Technology Services



INDIANA UNIVERSITY

Download Cytoscape

www.cytoscape.org



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Follow Along

<https://iu.box.com/v/network-graphs2017Fa>



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY

University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Digital Arts and Humanities Workshop Series – Fall 2017

Fridays @ noon -- Scholars Commons IQ-Wall

Date	Topic	Presenter
Aug. 25	Intro to Visualization	Michael Boyles
Sep. 1	Intro to Digital Humanities	Tassie Gniady
Sep. 8	Virtual Reality	Bill Sherman
Sep. 15	Intro to R	Tassie Gniady
Sep. 22	Advanced Media	Chris Eller
Sep. 29	Augmented Reality	Chauncey Frend
Oct. 13	R for Text	Tassie Gniady
Oct. 20	Network Graphs	David Kloster
Oct. 27	R for Twitter	Tassie Gniady
Nov. 3	3D Scanning and Printing	Jeff Rogers
Nov. 10	3D Photogrammetry	Tassie Gniady
Dec. 1	IQ-Tables & Touch-Enabled Software Workflows	David Reagan



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services

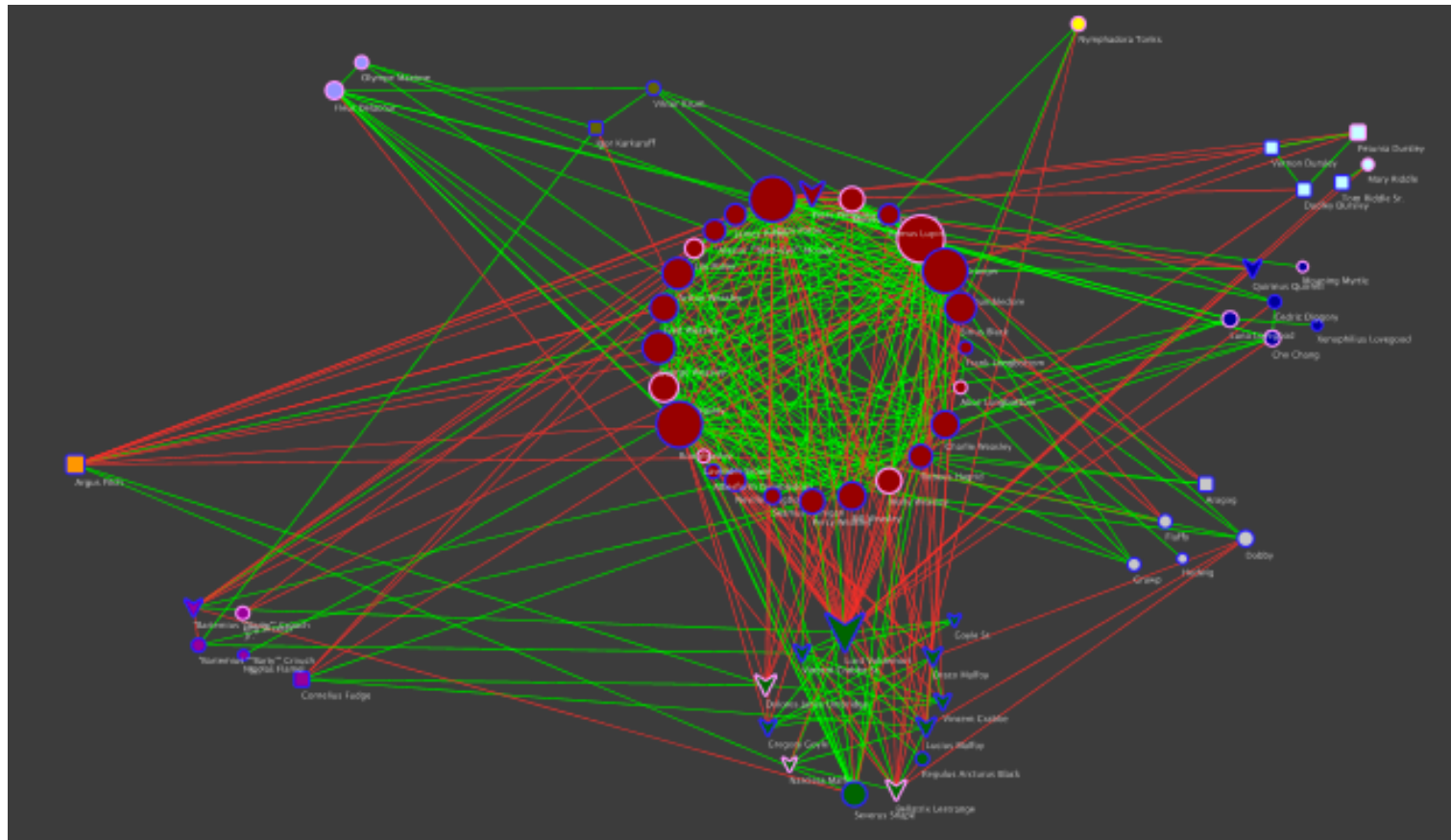


**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



What is a Network Graph and Why Make One?

RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



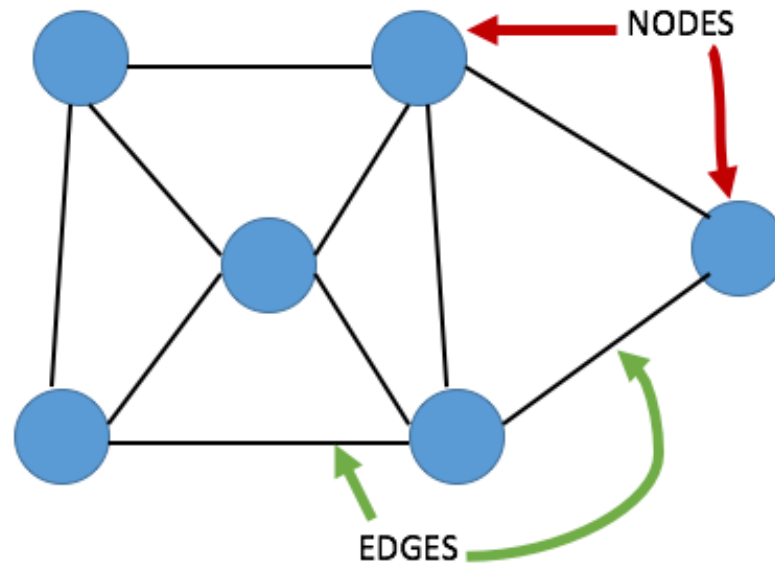
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Network Graphs...The Basics

- Nodes (vertices) = Objects
- Edges = The relationship(s)



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

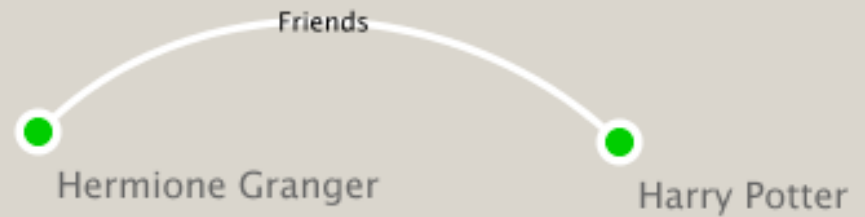


Terms to Know

Directed



Undirected



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services

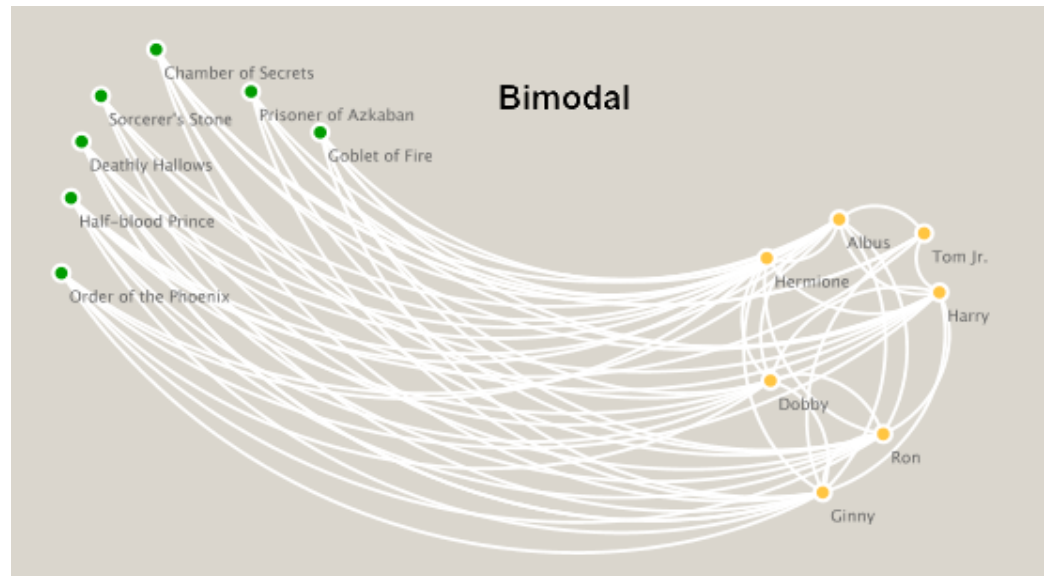
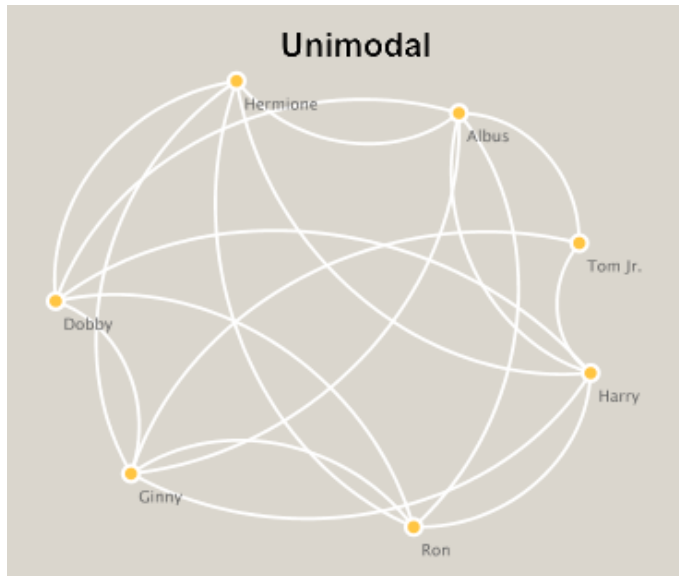


**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Types of Networks



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services

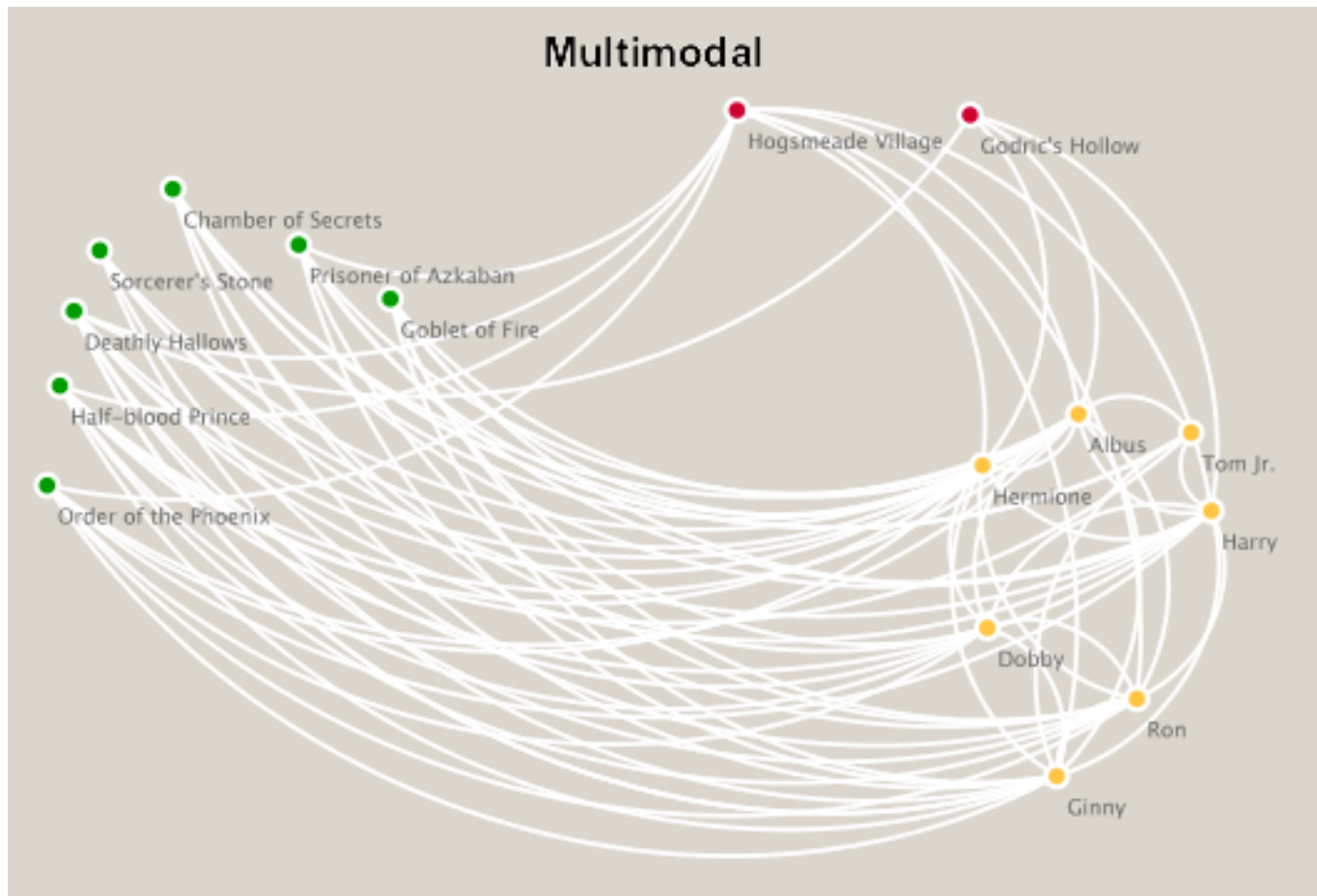


**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Types of Networks



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



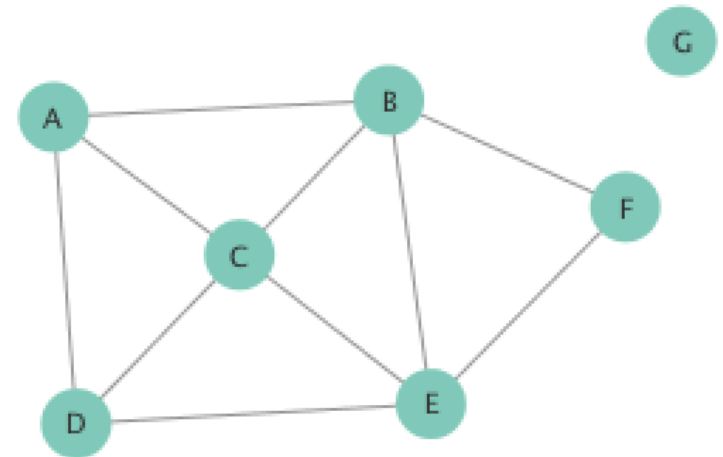
**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Node Attributes

- Degree - number of edges that connect to a node. For example, **Node C** has a degree of four and **Node F** only has a degree of 2.
- Isolates – Nodes not connected to any others edges. **Node G** is an isolate.



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



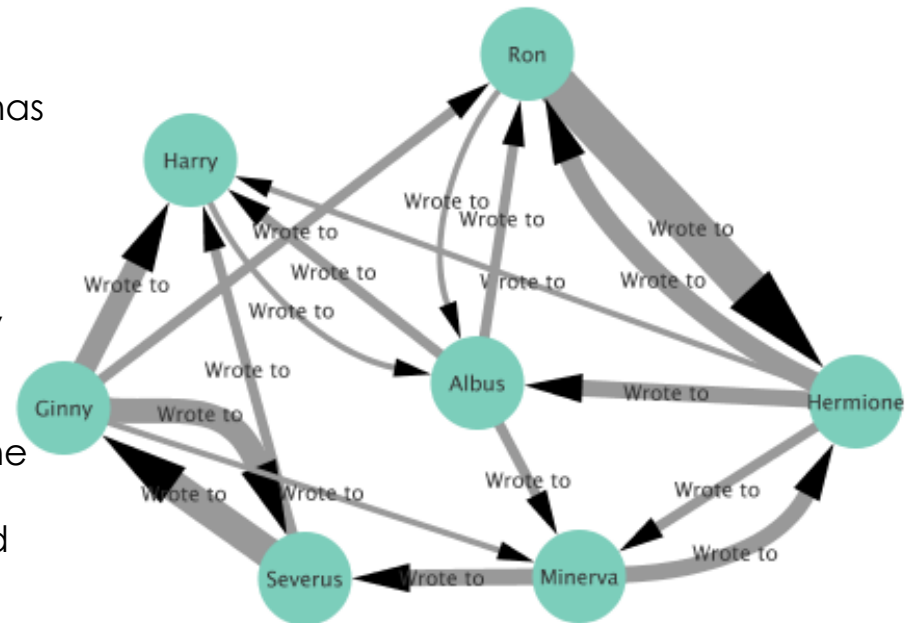
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Attributes Cont...

- In-degree - is calculated by determining the number of edges that point to a node
 - **Ron** has an in-degree of 3 and **Ginny** only has an in-degree of 1.
- Out-degree – is calculated by determining the number of edges that point away from a node
 - **Hermione** has an out-degree of 4 and **Harry** has an out-degree of 1.
- Weight (edge attribute) – Strength of the tie by the thickness of the edges between nodes
 - In the example, the edge between **Ron** and **Hermione** is Strongest (Wrote most letters).



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services








PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Popular Network Software

SOFTWARE	Ease of Use	Pros	Cons
 Cytoscape	<ul style="list-style-type: none"> No programming knowledge required Minimal understanding of networks Basic excel knowledge 	<ul style="list-style-type: none"> Scales with data User friendly Lots of options Well supported 	<ul style="list-style-type: none"> Made for hard sciences
 igraph	<ul style="list-style-type: none"> Requires use of R, Python, or C 	<ul style="list-style-type: none"> Can be used with multiple programming languages Allows for more manipulation of network aesthetics 	<ul style="list-style-type: none"> Requires some programming knowledge (we can help with that)
 Gephi	<ul style="list-style-type: none"> Intermediate knowledge of excel needed (ex. VLOOKUP) 	<ul style="list-style-type: none"> Allows for aesthetically pleasing networks Intuitive GUI 	<ul style="list-style-type: none"> Not supported well (4 years between updates) Still has reported issues with Macs
 Palladio	<ul style="list-style-type: none"> Easy to follow tutorials on the web page No programming knowledge required 	<ul style="list-style-type: none"> More options User friendly instructions 	<ul style="list-style-type: none"> Mostly intended for geospatial networks
 Fusion Tables	<ul style="list-style-type: none"> Uses .csv, .tsv, .txt, .kml or google sheets 	<ul style="list-style-type: none"> Basic Easy to use 	<ul style="list-style-type: none"> Minimal options



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Info and Examples on Networks in the Humanities



Scott Weingart

[Demystifying Networks Parts I & II.](#)
[Journal of Digital Humanities.](#)



Miriam Posner

- [Cytoscape Tutorial](#)
- [Blog post about Cytoscape](#)



Rachael Cohen & Angie Thorpe

[IU Bloomington and Kokomo library](#)
[Search terms by subject](#)



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Download Cytoscape

www.cytoscape.org



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Creating a Network

(General Steps)

- Determine your problem/scholarly question
- Assess what data you have and what is available
- Make analog network graph on paper
- Organize and clean your data (determine what your objects are and the nature of their relationship)
- Plug your data into your software (in our case Cytoscape)
- Map your network graph (assign node colors, node size, edge weights, etc...)
- Analyze and decide if the network solves your problem/answers your question

NOTE: Should practice on something you know really well first. (So you can see if your results are making sense)



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



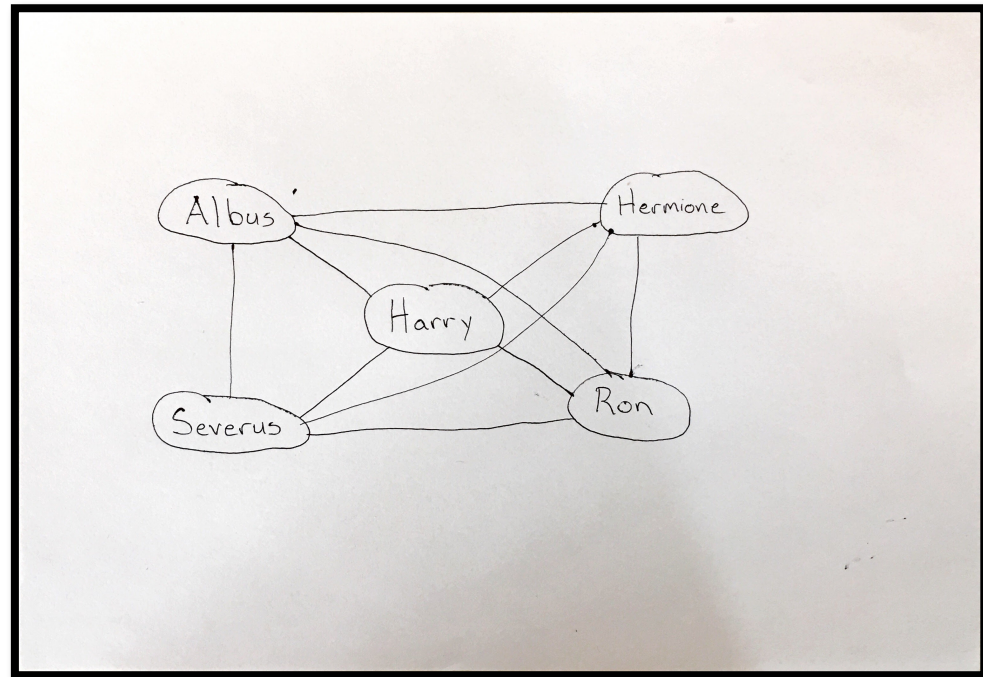
**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Network Creation using Cytoscape

- Steps 1 & 2 already done
- Make analog network graph



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Edge and Node List

- Organize
 - In a spreadsheet:
 - Create an edge list (determines objects, their relationship to one another, and edge attributes)
 - Create a node list (determines object attributes)

(NOTE: Make sure spelling and capitalization is exact and universal in both lists)

Normally, this would be the next step, however...

Edge

Source	Target	Type
Regulus Arcturus Black	Sirius Black	
Regulus Arcturus Black	Bellatrix Lestrage	
Regulus Arcturus Black	Lord Voldemort	
Sirius Black	Regulus Arcturus Black	
Sirius Black	Albus Dumbledore	
Sirius Black	Hermione Granger	
Sirius Black	Bellatrix Lestrage	
Sirius Black	Remus Lupin	
Sirius Black	Lucius Malfoy	
Sirius Black	Narcissa Malfoy	
Sirius Black	Minerva McGonagall	
Sirius Black	Alastor "Mad-Eye" Moody	
Sirius Black	Peter Pettigrew	
Sirius Black	Harry Potter	
Sirius Black	James Potter	
Sirius Black	Lily Potter	
Sirius Black	Lord Voldemort	
Sirius Black	Severus Snape	
Sirius Black	Nymphadora Tonks	
Sirius Black	Arthur Weasley	
Sirius Black	Fred Weasley	
Sirius Black	George Weasley	
Sirius Black	Ginny Weasley	
Sirius Black	Ron Weasley	
Sirius Black	Hermione Granger	
Sirius Black	Neville Longbottom	
Sirius Black	Ron Weasley	

Node

Character	Alignment	Gender	House
Regulus Arcturus Black	Good	Male	Slytherin
Sirius Black	Good	Male	Gryffindor
Lavender Brown	Good	Female	Gryffindor
Cho Chang	Good	Female	Ravenclaw
Vincent Crabbe	Evil	Male	Slytherin
Vincent Crabbe	Evil	Male	Slytherin
Bartemius "B" Croaker	Good	Male	Unknown
Bartemius "B" Croaker	Evil	Male	Unknown
Fleur Delacour	Good	Female	Beauxbatons
Cedric Diggory	Good	Male	Ravenclaw
Aberforth Dumbledore	Good	Male	Gryffindor



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Download Google Sheets Sample

To save time, download this edge list here:

<http://bit.ly/edgelisthp>

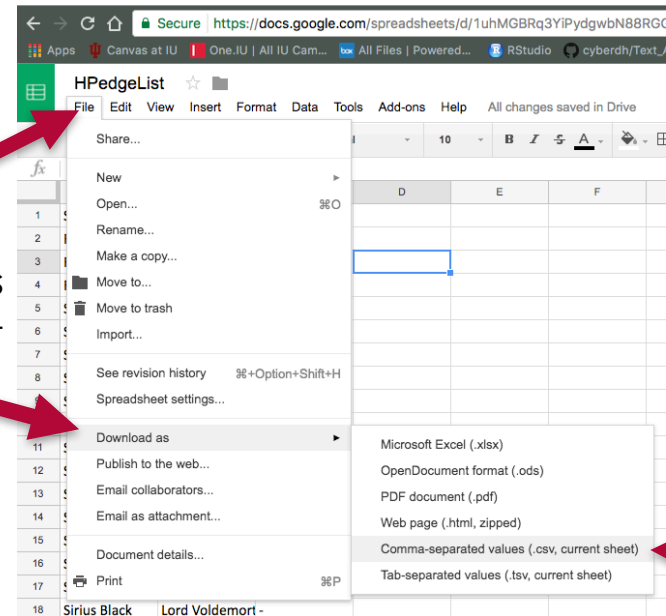
And this node list here:

<http://bit.ly/hpnodeлист>

Go to File>Download as>Comma-separated values (.csv, current sheet) for both the edge and node list

Save as HPedgeList.csv and HPnodeList.csv
Respectively

NOTE: Better to download as .csv



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services




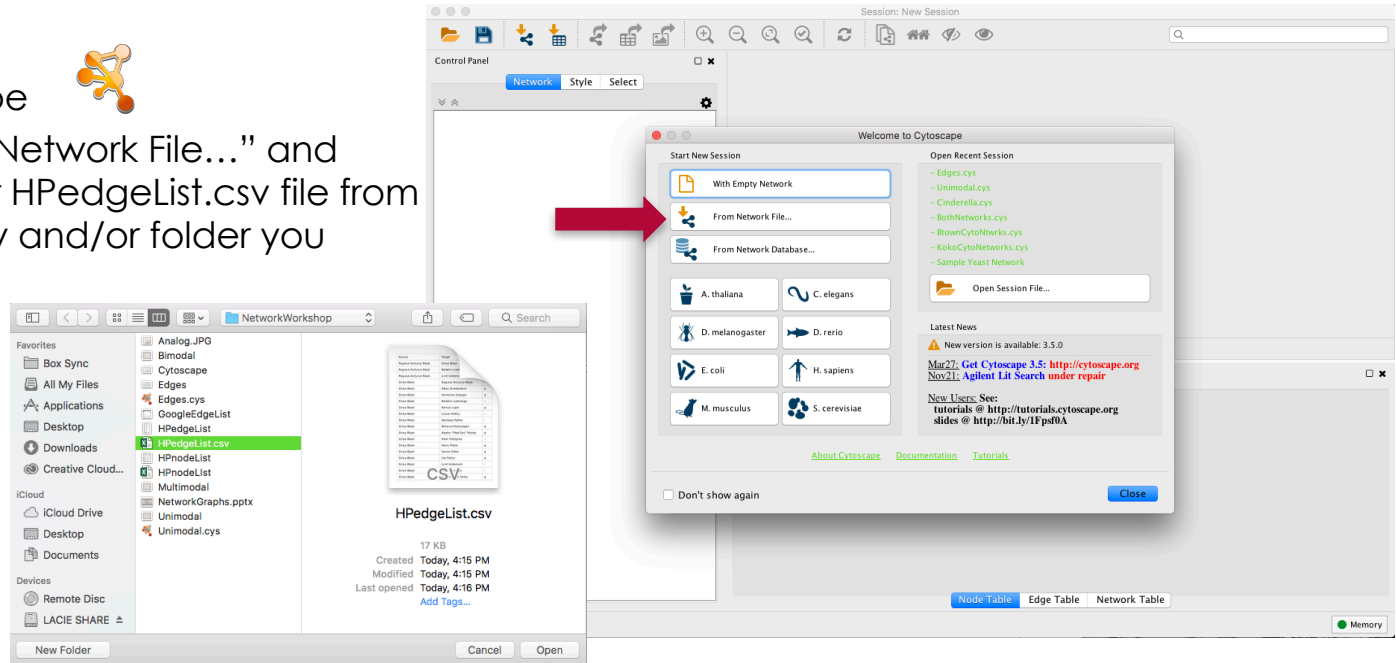
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Upload .csv Edge List

- Open Cytoscape
 - Click "From Network File..." and choose your HPedgeList.csv file from the directory and/or folder you saved it in



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



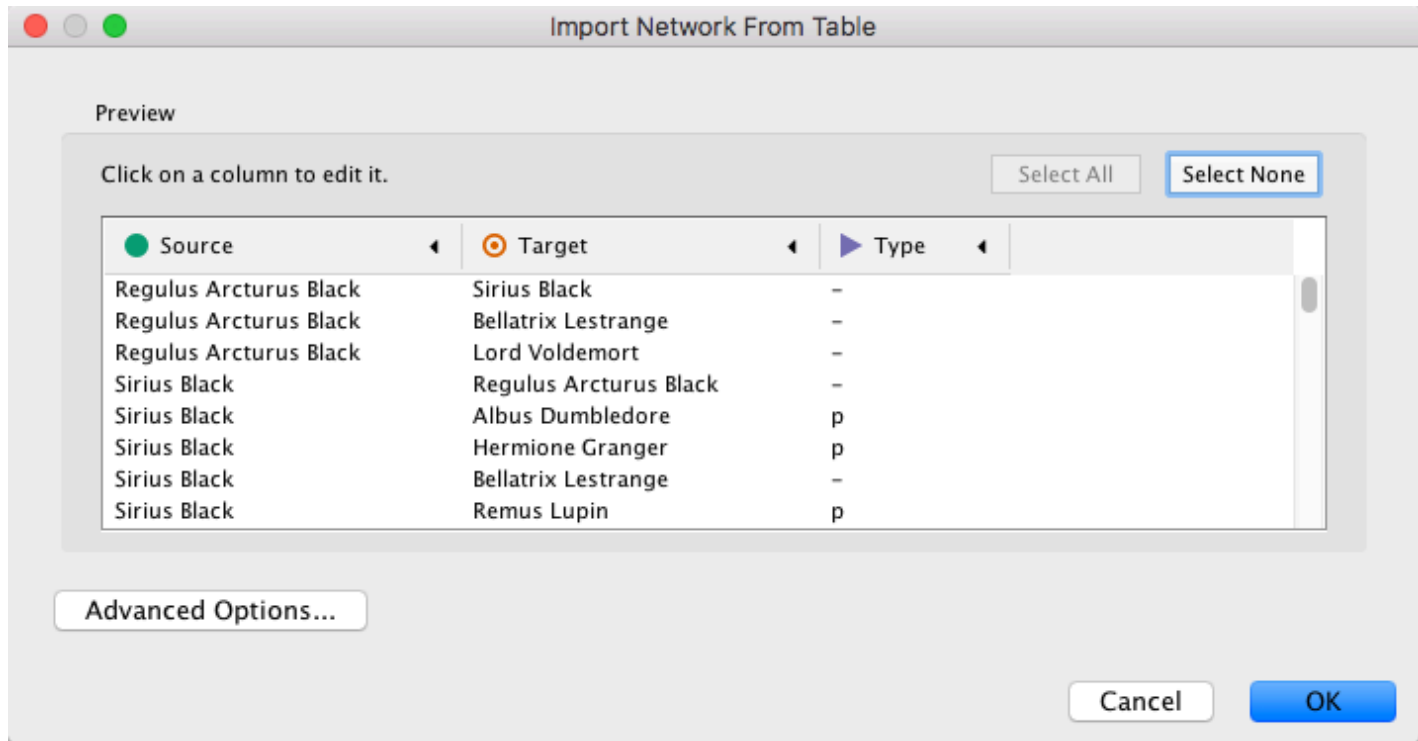
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Source and Target

- Make sure the “source” column and “target” column are designated as such then click “ok”



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



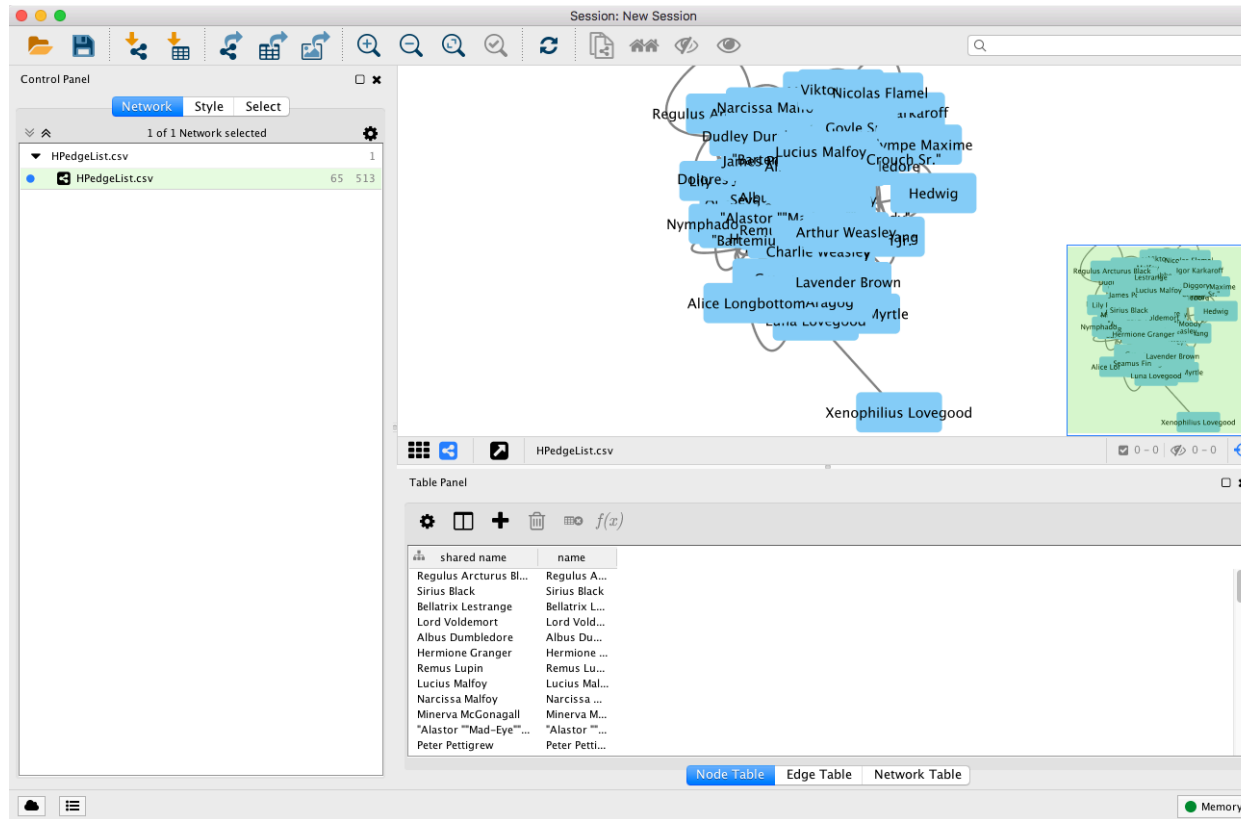
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Hot Mess

- You should now have what appears to be a jumbled mess of spaghetti and meatballs.



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



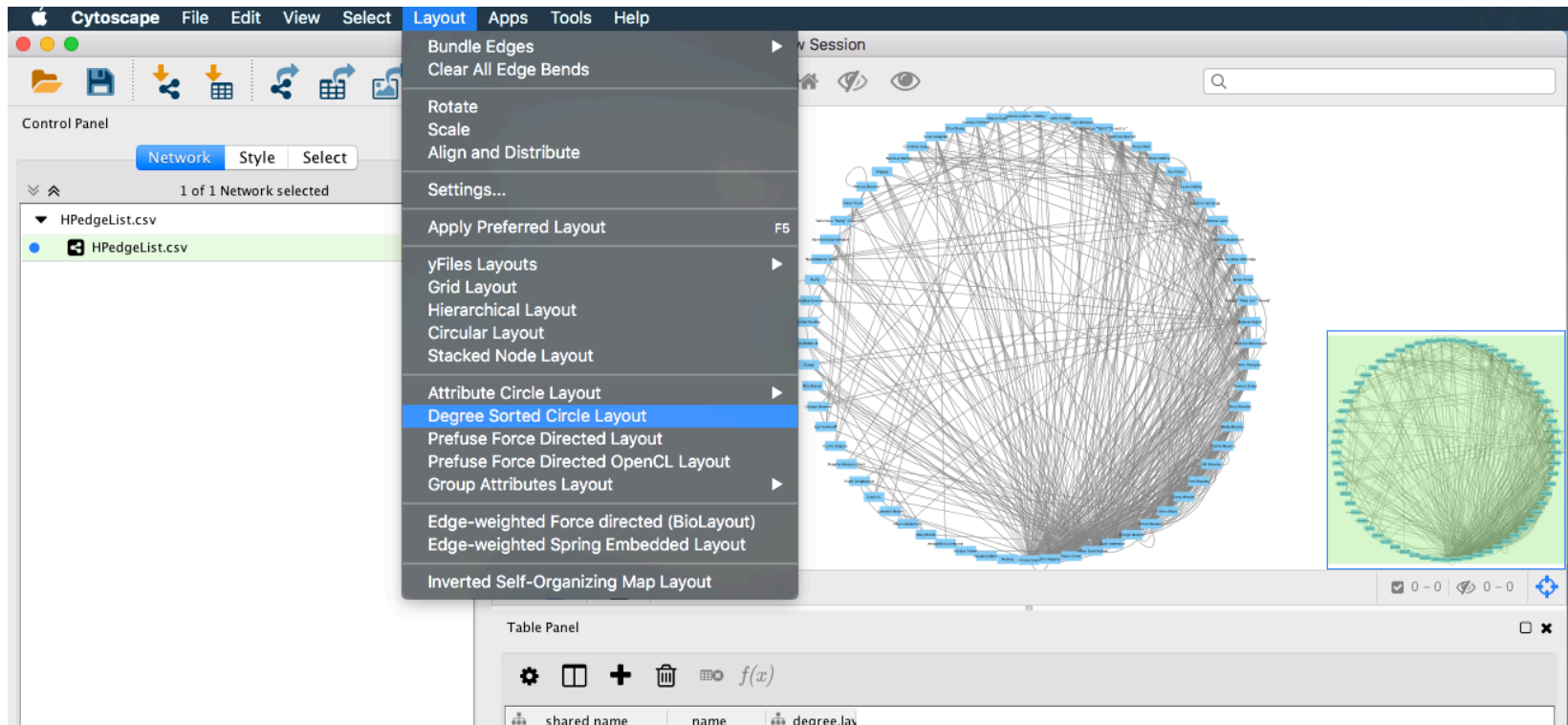
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Layout

- Go to Layout>Degree Sorted Circle Layout (or one that you prefer)



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Style

- This is “Sample 3” or feel free to choose one of your own

The screenshot shows the Cytoscape software interface. On the left, the 'Control Panel' is open to the 'Style' tab, displaying a grid of network styles. The 'Sample3' style is selected, which features a blue circular node and a blue directed edge. The main window displays a large, dense network graph with many nodes and edges. A smaller inset window shows a zoomed-in view of the network. Below the main window, a table titled 'HPedgeList.csv' is visible, showing a list of nodes and their degrees.

name	degree
Regulus A...	5
Sirius Black	33
Bellatrix L...	16
Lord Vold...	46
Albus Du...	55
Hermione ...	59
Remus Lu...	17
Lucius Mal...	15
Narcissa M...	7
Minerva M...	23
"Alastor "Mad-Eye" ...	19
Peter Pettigrew	23



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



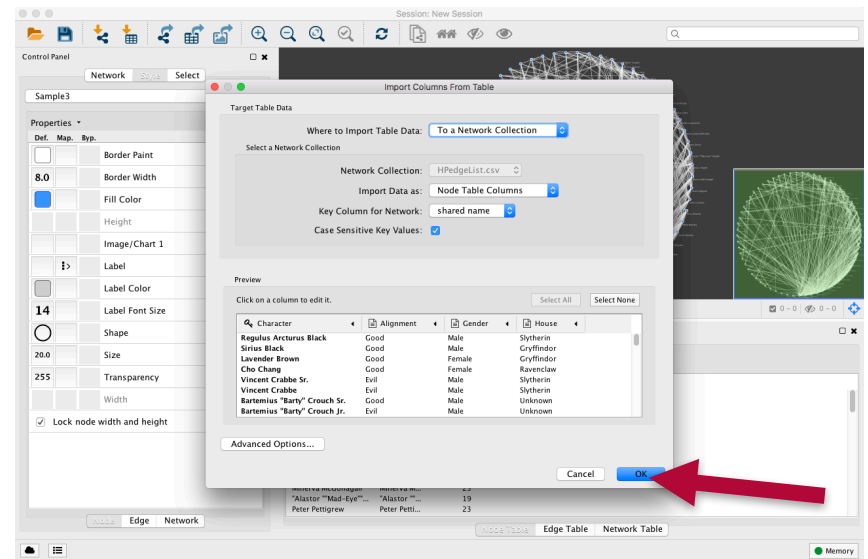
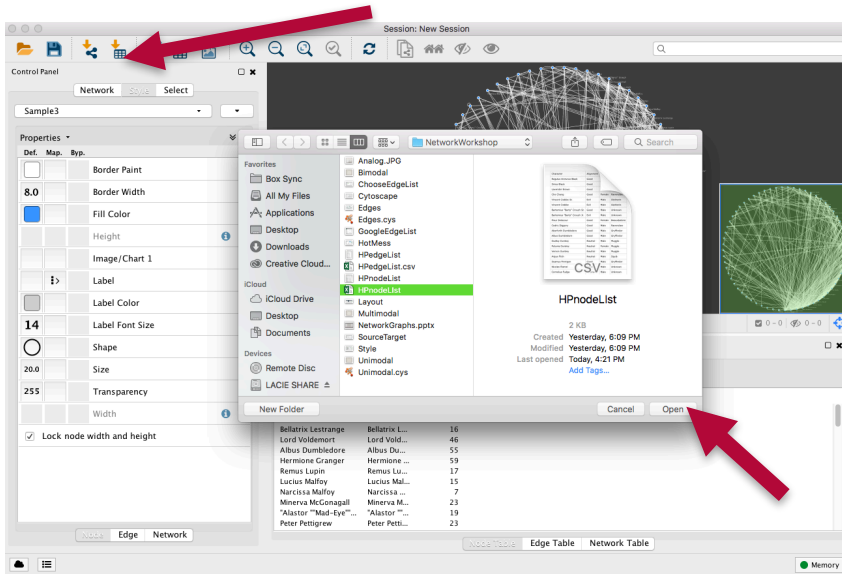
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Upload Node List

- Click on the “Import Table From File” icon, open file, then click “ok”



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services



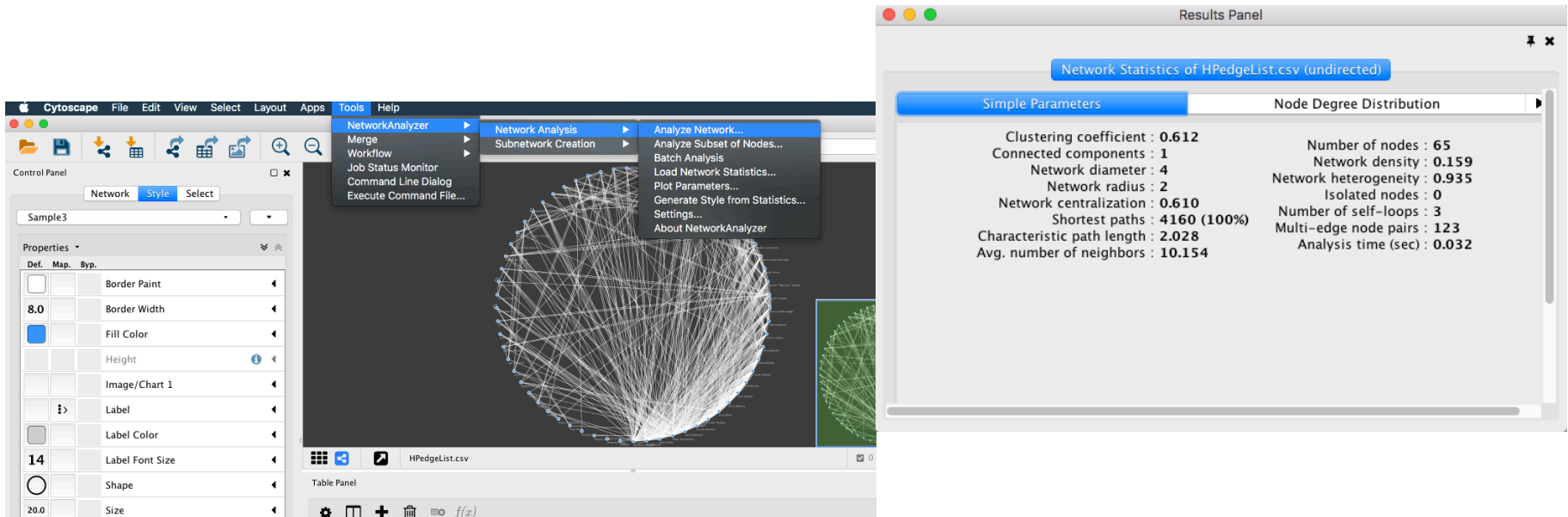
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Analyze Network

- Go to Tools>NetworkAnalyzer>Network Analysis>Analyze Network
- The “Results Panel” should appear. Look at some of the terms and numbers.



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



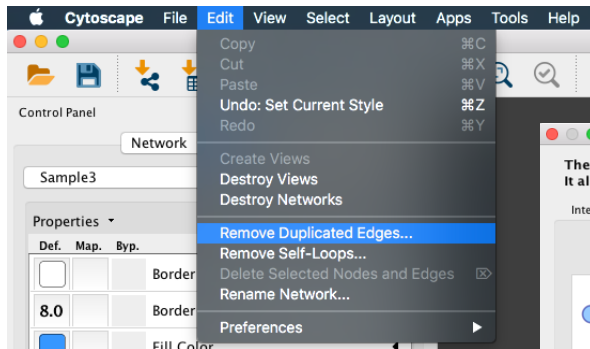
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY

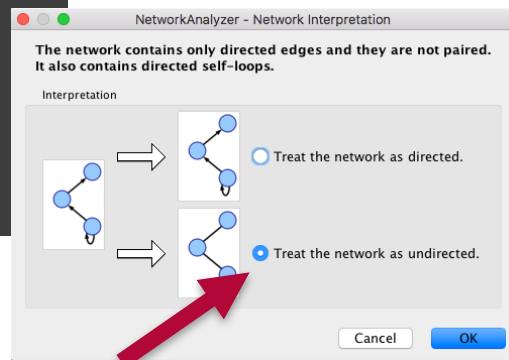


Remove Duplicates

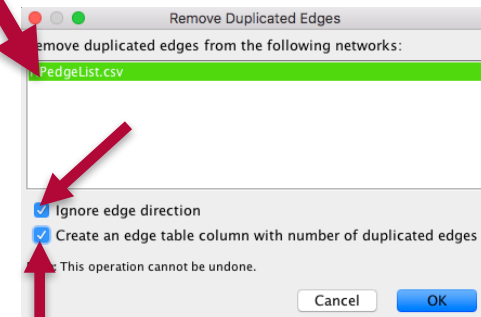
Go to Edit>Remove Duplicated Edges



Choose "Treat the network as undirected"



Make sure the file is highlighted and both boxes are checked



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



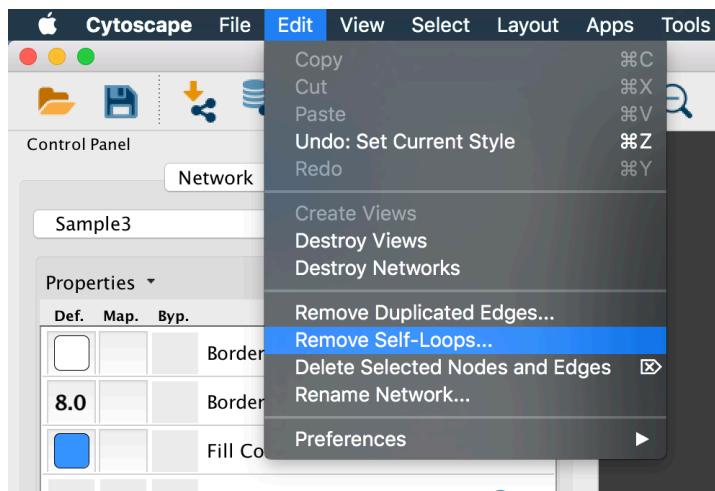
PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY

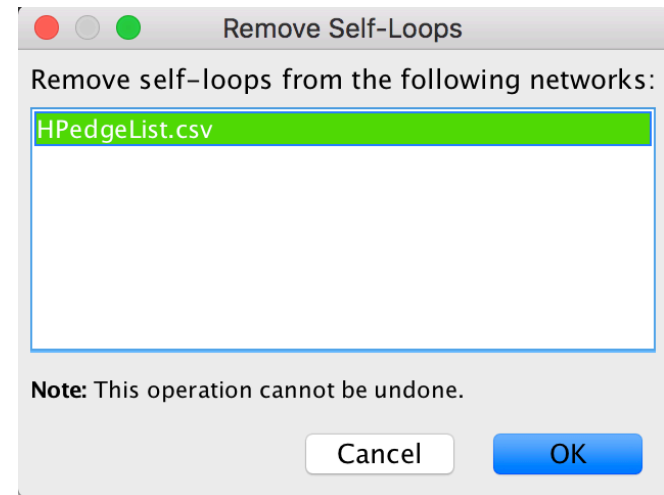


Remove Self-Loops

Go to Edit>Remove Self-Loops



Make sure file is highlighted
Then click "ok"



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services

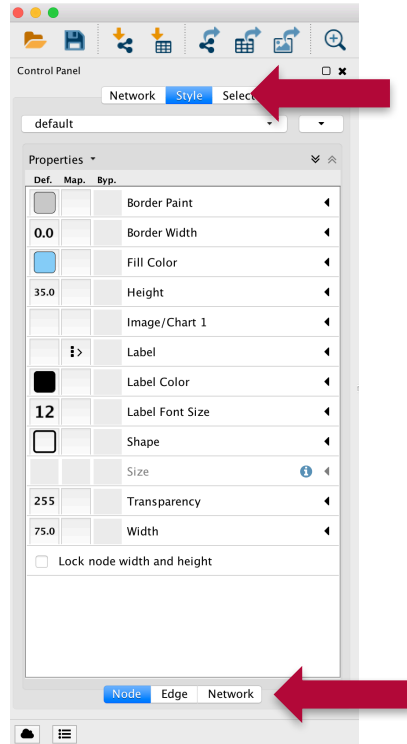


PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Let's Map



- To map the nodes and edges make sure you have “Style” selected on the top of the left panel
- Make sure you have the proper entity selected on the bottom as well. Either “Node” or “Edge”, depending on which one you are interested in mapping
- Now let's see what this network can reveal...

NOTE: Don't be afraid to play around with the software or even break it.



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Important Things to Note

- Other visualizations may better reflect or help display your research:
 - May not accurately reflect your research (may be misleading)
 - May not be the best way to visualize your findings (not aid in understanding)
- Be aware of what the software you are using does:
 - May be made for other purposes (may need to alter steps)
 - May work better with other types of data (Geospatial networks)
 - May require more tech skills than you have or have time to acquire (since you have us, and many others that can help with this, it is not as big of a problem here at IU)
- Creating a network is not the end:
 - Keep careful track of every step and decision in making your network
 - Analyze the network carefully
 - Don't be afraid to start over (will almost always need to at least once, especially if you have a large network)



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services

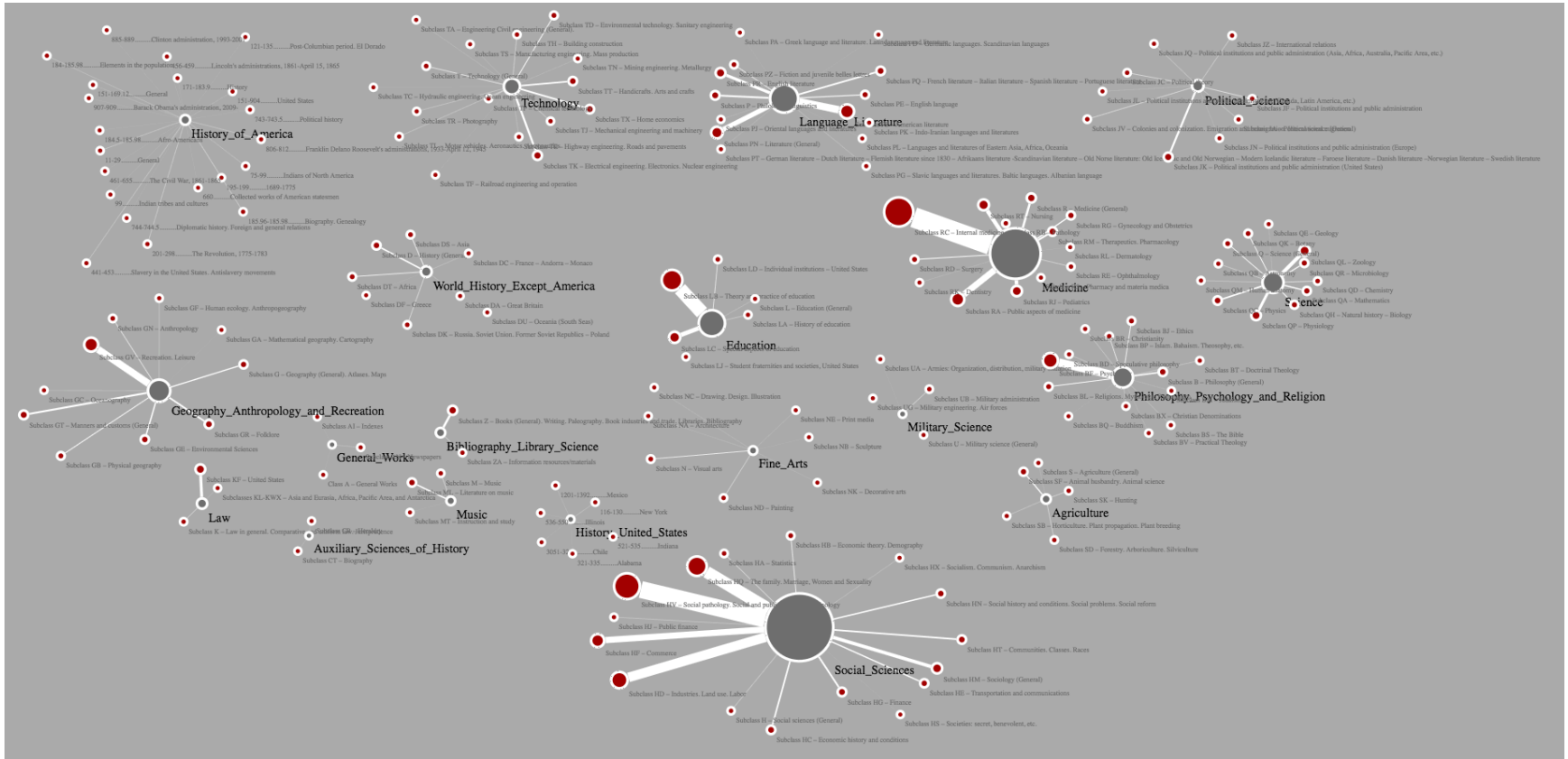


**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Sample Network:



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY

University Information Technology Services

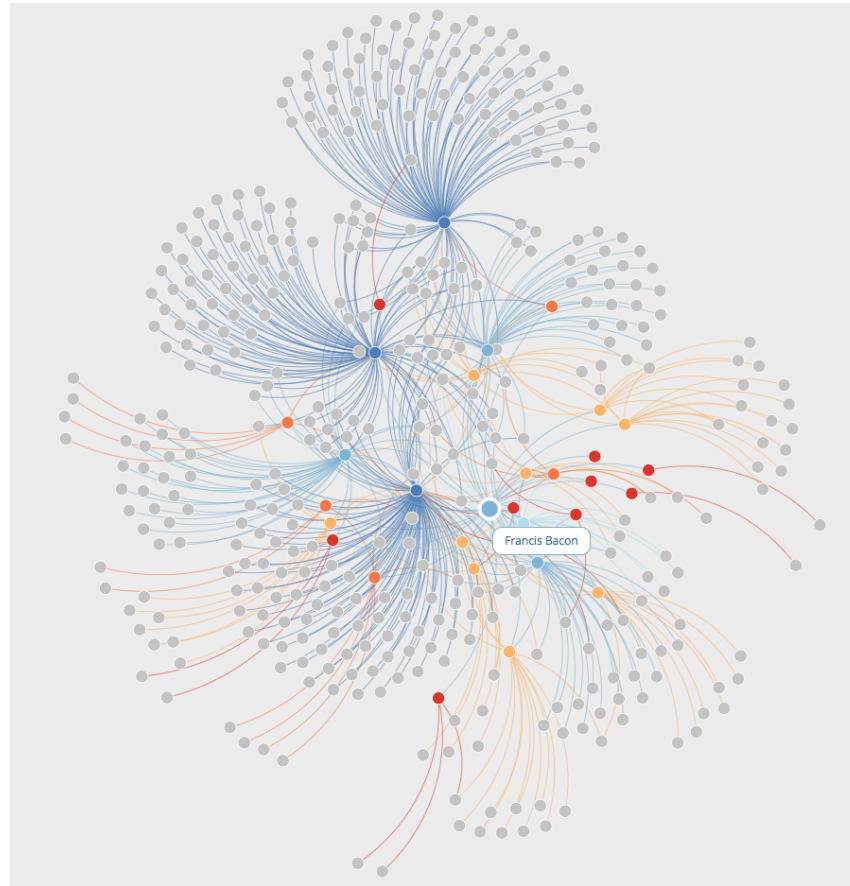


PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



Sample Network: Six Degrees of Francis Bacon



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY
University Information Technology Services

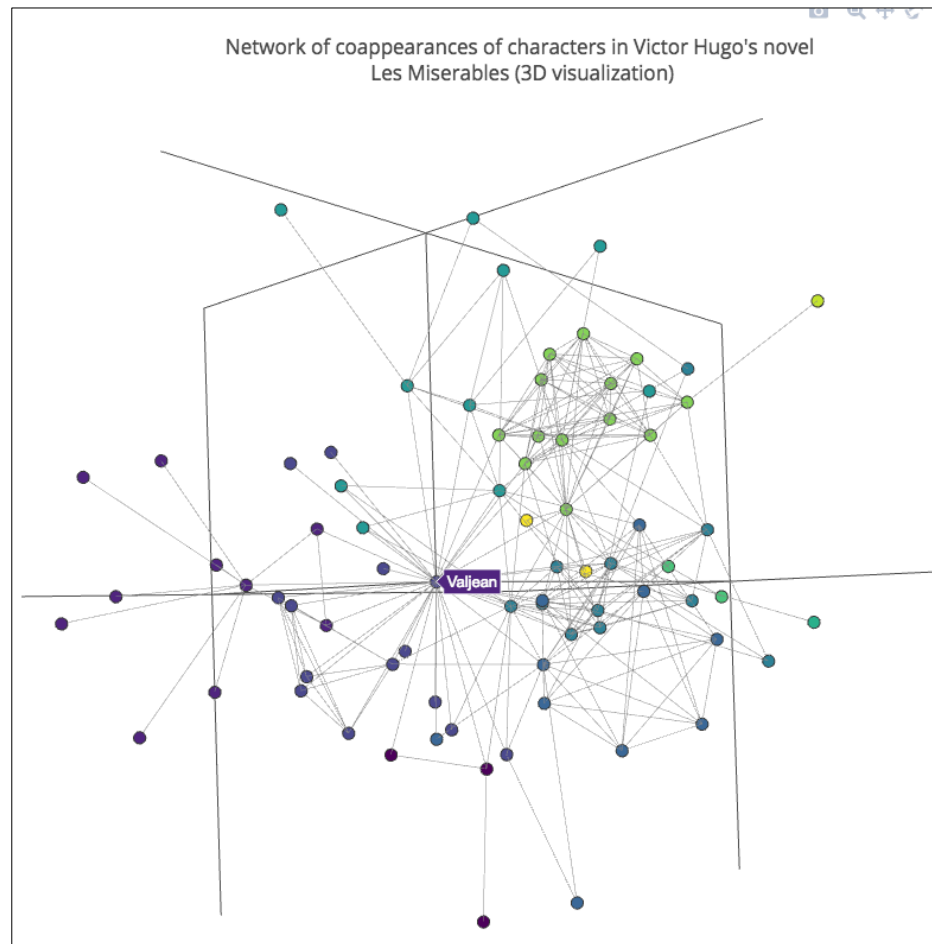


**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



Sample Network: iGraph and *Les Miserables*



RESEARCH
TECHNOLOGIES

INDIANA UNIVERSITY
University Information Technology Services



PERVASIVE TECHNOLOGY
INSTITUTE

INDIANA UNIVERSITY



More Information

The CyberDH team can answer Network Graph questions. Get in touch!

Contact:

David Kloster

Cyberinfrastructure for Digital Humanities Coordinator

cyberdh@iu.edu

[Cyber DH Blog](#)



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY

University Information Technology Services



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY

